AMERICAN
PUBLIC
TRANSPORTATION
ASSOCIATION

May 23, 2011

U.S. Department of Transportation
Docket Operations
M-30, West Building Ground Floor
Room W12-140
1200 New Jersey Avenue, SE
Washington, DC 20590

RE: Docket No. FRA-2009-0043, Notice No. 1

Dear Docket Clerk:

On behalf of the 1,500 member organizations of the American Public Transportation Association (APTA), I write to provide comments on the Federal Railroad Administration's (FRA) Notice of Proposed Rulemaking (NPRM) concerning Hours of Service of Railroad Employees; Substantive Regulations for Train Employees Providing Commuter and Intercity Rail Passenger Transportation; Conforming Amendments to Recordkeeping Requirements, announced March 22, 2011, at 76 FR 16200.

About APTA

APTA is a non-profit international trade association of 1,500 public and private member organizations, including public transit systems; high-speed intercity passenger rail agencies; planning, design, construction and finance firms; product and service providers; academic institutions; and state associations and departments of transportation. More than ninety percent of Americans who use public transportation are served by APTA member transit systems.

Our Comments

Validation and Calibration of Fatigue Models

The Fatigue Avoidance and Scheduling Tool (FAST) sponsored by FRA was tested and calibrated based on freight railroad accident data. It was never independently validated for passenger railroad service "because there were not enough accidents in the relevant time period to obtain statistically significant results." 76 FR 16200, 16207. Despite a lack of data, FRA sought to extrapolate their data to passenger service by assuming "the tasks associated with freight and passenger train operations are actually highly similar." Id. We disagree.

Many tasks are different in passenger operations and the operating environment those tasks are conducted in is substantially different. Cab signaling, civil speed enforcement, and other factors all contribute to the dissimilarity of the tasks. Passenger railroad conductors spend the majority of their time in tasks related to fare collection, passenger information, and customer care, and rarely are called upon to couple or uncouple cars, throw switches, or other such tasks common to freight operations.

Given the limited data and questionable link between freight railroad accidents and day-to-day passenger operations, we believe FRA should undertake a full review of its FAST validation, specifically looking to the full range of operations rather than the limited accident data available, and specifically looking at the unique nature of passenger operations.

Understanding this review would otherwise cause FRA to miss Congressionally-mandated time limits, we believe FRA should provisionally move forward with a FAST threshold of 70, as proposed, and agree to conduct the review expeditiously thereafter. Moreover, we believe FRA should agree to subject this review to independent peer review prior updating the threshold.

Calibration of Fatigue Models

In the NPRM, FRA described the process it employed to test its FAST. That process led FRA to determine a proper fatigue threshold under that model was 70. FRA further accepted the validity of the Fatigue Audit InterDyne Model (FAID) and undertook to determine an appropriate, approximately equivalent fatigue threshold under FAID. We object to FRA's calibration efforts that resulted in its proposal to set the fatigue threshold under FAID at 60.

The proposed calibration equating a FAST score of 70 with a FAID score of 60 results in a significant number of illogical results. Many schedules successfully meeting the threshold under FAST fail under FAID where the threshold score is set at 60. Moreover, we note that although FRA stated the NPRM is "consistent with the recommendation of the Working Group (WG) and the full" Rail Safety Advisory Committee (RSAC), the Working Group, in fact, conducted its analysis and discussions based on a FAID threshold of 90. FRA's decision to propose a FAID threshold of 60 post-dated the RSAC vote.

As demonstrated in the work of Dr. Patrick Sherry, whose report is attached, the strongest, most reliable comparative value to equate to FAST 70 is FAID 90. Even at 90, the FAID system yields a slightly more conservative result than FAST. In fact, with that threshold score of 90, almost two percent of schedules that meet FAST 70 will fail under FAID and prompt railroads to mitigate the effects of fatigue for workers under those schedules. We strongly encourage FRA to adopt 90 as the fatigue threshold under FAID.

Economic Analysis

The economic impact of the proposed rule will be substantially higher than estimated by FRA, even if FRA adopts our recommendation to adopt FAID 90 as the threshold value under that model.

During the course of the RSAC Working Group discussions leading to this NPRM, a sample of APTA member railroads had some or all of their schedules modeled. That modeling demonstrated that commuter railroads will be required to hire and train new engineers and conductors or cut service specifically to comply with the proposed rule. Estimates from just five commuter railroads demonstrate the annual costs of additional personnel based on this proposed rule will exceed $15M for just those five railroads. Adding just those five railroads to the extra personnel costs Amtrak would expect pushes the annual personnel costs of compliance over $27M per year.

Fatigue training costs will likewise dwarf FRA estimates. Those same five railroads have determined annual fatigue training is likely to add over $1.8M to their costs.

Moreover, it appears that licensing costs for FAST software would far exceed FRA estimates. The FAST model, whose development was sponsored by FRA and initially represented to be made available to railroads at no cost would, in fact cost approximately $500,000 for a single railroad. By contrast, the FAID model would be available at about five percent of the cost of FAST.

Shared Extra Board

While the NPRM addresses the use of freight employees as pilots, it does not address routine employment of a shared extra board between freight and passenger operations. The preponderance of work for these employees remains freight, even if they are used for occasional passenger work. Since it would be nearly impossible to interface the freight and passenger standards to determine whether these shared extra board schedules met the threshold values, and the standards applicable to freight employees are more restrictive than those related to passenger operations, FRA should simply evaluate shared extra board schedules under freight standards.

Parameters of Type 1 Schedules

FRA has proposed classifying all schedules falling entirely in the 4 a.m. to 8 p.m. time frame as Type 1, essentially requiring no additional modeling or analysis. We propose those parameters be expanded to 3 a.m. to 8 p.m. This would be consistent with FRA’s statement that the “risk of a human factor accident is increased by 20 percent by working during the hours from midnight to 3 a.m.” 76 FR 16200, 16207. This would link the Type 1 definition to FRA’s scientific research.

Treatment of Service Disruptions

FRA has proposed to treat delays based on casualties, accidents, acts of God, and unknown and unforeseeable causes as outside the application of the Subpart. Proposed section 228.403(a), 76 FR 16200, 16226. We believe FRA should affirmatively state that mechanical breakdowns, signal failures, switch failures, or other such conditions fall within the definition of “cause[s] unknown and unforeseeable” for purposes of that section. The effects of these delays on fatigue are no different from any of the other listed reasons and it would be impractical to compromise an approved schedule for what would most likely be a *de minimis* impact.

Clarifications

There are a number of issues FRA should clarify in publishing its final rule. These include the following:

FRA should specifically note that schedule analysis and approval may be shared among railroads. A schedule analyzed and approved for railroad A may also be used by railroad B without additional analysis. This will ease the economic burden of modeling without compromising safety and may well allow smaller railroads to forego the extensive costs of obtaining modeling software licenses and conducting analysis.

FRA should clarify that railroads are not required to immediately implement the proposed limitations on consecutive calendar days worked (proposed section 228.405) on the effective date of the final rule. Railroads will have to complete their schedule analysis before determining Type 1 and Type 2 assignments and confer with affected employees prior to implementation. Implementation should be explicitly stated as 180 days after the effective date of the regulation.

FRA should provide substantive examples, and questions and answers to explain the consecutive day restrictions in the rule. It has become abundantly clear that even those involved in the RSAC process do not necessarily understand FRA's intent in this respect.

FRA should specifically note that railroads need not wait for FRA's written approval of submittals prior to initiating their implementation strategies, particularly concerning employee training, mitigation plans, and other actions requiring significant implementation periods.

We appreciate the opportunity to assist FRA in this important endeavor. For additional information, please contact James LaRusch, APTA's chief counsel and vice president - corporate affairs, at (202) 496-4808 or jlarusch@apta.com.

Sincerely yours,

William Millar

William Millar
President

WM/jpl

PRELIMINARY RESULTS OF FAST- FAID CALIBRATION STUDY

Patrick Sherry, PhD
University of Denver

Prepared
By
Dr. Patrick Sherry

University of Denver
National Center for Intermodal Transportation

May 23, 2011

Executive Summary

There has been an increased interest in the use of bio mathematical models to understand and predict the impact of extended work hours, exceptional duty rosters, and other work related demands. The present study sought to improve on previously published analyses by analyzing a more representative sample of work schedules typical of the everyday operations of the commuter rail or intercity passenger rail industry. A representative sample of work schedules was obtained that consisted of 101 work schedules in which 61% were morning starts, 36% were afternoon, and 3% were midnight starts. These schedules were then analyzed to produce FAST scores for every 30 minute interval that the employee was working. The results indicate the presence of a highly statistically significant relationship between the two models which supports the assumption that the two models are measuring similar phenomena. Therefore we can assume that the FAID model is also validated and that cutoff scores on the FAID model can also be equated to cutoffs on the previously validated FAST model. The exact score conversion between FAST and FAID is presented using the linear conversion model.

Table of Contents

Introduction

There has been an increased interest in the use of bio mathematical models to understand and predict the impact of extended work hours, exceptional duty rosters, and other work related demands. Bio mathematical models have been developed in various laboratories around the world with the intention of modeling and predicting the physiological and cognitive responses to a variety of different conditions to which the individual has been exposed.

The accuracy of these models for both describing and predicting human behavior was the subject of a conference on fatigue modeling held in Seattle Washington in 2002 and described in a special of issue *Aviation, Space and Environmental Medicine* (Neri, 2004). The most popular and well published models were described and compared using five separate sets of data that were thought to represent common and extreme conditions in the aviation and railroad industry. The conference organizers asked the authors of the models to utilize the prepared data sets and to analyze the data using their models. The models were then compared to determine how well they accounted for the data that they were attempting to model. The results of the conference indicated that none of the models was much better than any of the others in accounting for and predicting human fatigue. In fact the results of the analyses comparing the various models concluded that none of the models was very different from any other. In addition, overall none of the modes was very good at explain or predicting the restricted sleep scenario conditions, the kind of sleep schedule typically faced by people in the rail industry.

Model Calibration

In its November 2010 report, "Procedures *for Validation and Calibration of Human Fatigue Models: The Fatigue Audit InterDyne Tool,*" the Federal Railroad Administration (FRA) described a method for validating fatigue models that involved demonstrating a statistically significant relationship with an already or previously validated model. Previously the FAST model (Hursh, et al., 2004) had been related to an increased risk of human factor caused accidents with scores on the FAST model below 70 (Hursh, Raslear, Kaye, and Fanzone (2006). In the Tabac & Raslear (2010) study, a significant relationship between FAST and FAID was demonstrated and a calibration set as well.

Results of this analysis demonstrated that there was a significant linear relationship between the FAST and the FAID scores and that a biomathematical model was able to be determined. In fact, the published correlation coefficient between FAST and FAID scores was >.90. However, this relationship was based on bin, or ten point interval, means of the FAST scores, comprised of the scores that fell within a ten point range of FAST scores rather than individual pairs of scores.

Such an approach reduces the normal variation in the relationship between the independent and dependent variables examined in this analysis. The application of linear regression techniques is typically undertaken with the assumption that the underlying distribution has a moderate amount of variability. By limiting the analyses to bin means the variability is thereby reduced and predictive and explanatory power is reduced considerably. A more robust application of linear regression requires the use of data with more variability.

Results of the Tabac & Raslear (2010) study determined that a FAID cutoff score of 60 corresponded to a FAST score of 70 following the linear transformation of the FAID score using the parameter weights and constants identified in the study. However, the identification of 60 as the corresponding equivalent to the FAST score may also be the result of unique characteristics of the data set used to generate the linear transformation equation. The data set identified consisted of work schedules of employees involved in either human factor or non human factor caused accidents in the freight industry. Inspection of the data provided by Hursh, Raslear, Kaye, and Fanzone (2006) reveals that most of the accident data provided fell 21:00 and 05:00 hours. Thus, this particular data set might have a slight bias towards lower levels of alertness and higher levels of fatigue. While such a data set is useful in showing the relationship between accident data and fatigue models it is not optimal for calibrating one model to another because the mean of the data set is weighted towards the fatigued end. In this there will likely be a preponderance of scores from both the FAST and the FAID model that would be in the range suggesting a higher risk for fatigue. These scores, due to the law of central tendency, would have the effect of skewing the distribution towards the fatigued end.

Since one goal of these studies is to provide a tool that can apply generally to the passenger rail industry, an alternative methodology would be to use a sample of typical work schedules drawn from the passenger rail industry. Moreover, since the goal is to establish a mathematical relationship between the two models a more robust relationship may be demonstrated by choosing a typical sample of work schedules that represent the likely activities of everyday operations. Thus, the present study sought to analyze a more representative sample of work schedules typical of the everyday operations of the commuter rail or intercity passenger rail industry.

Present Study

Based on the proposed alternative methodology for determining the best calibration of FAST and FAID it was proposed that a representative sample of schedules be analyzed according to the percentage of morning afternoon and midnight schedules. The data submitted suggested that some of the largest railroads had the following percentage breakdown of work schedules.

Percentage of Morning, Afternoon and Nighttime Schedules In Passenger Railroad Operations

On Duty		Off Duty		Metra	SEPTA	MNR	LIRR
3:30 AM	10:00 AM	11:30 AM	10:00 PM	65%	60%	62%	57%
10:00 AM	9:00 PM	1:00 PM	3:00 AM	32%	37%	32%	42%
9:30 PM	3:30 AM	7:00 AM	9:30 AM	2%	2%	6%	1%

Given that this percentage breakdown is consistent for four major commuter railroads a representative sample of work schedules was obtained that consisted of 101 work schedules. In this sample 61% were morning starts, 36% were afternoon, and 3% were midnight starts. These schedules were then analyzed to produce FAST scores for every 30 minute interval that the employee was working. Similarly, *InterDynamics* in Australia, publishers of FAID, analyzed the same data set and prepared a similar of set FAID scores during work periods for every 30 minute interval worked. The data for a typical schedule (e.g. schedule #240) was arranged as follows:

Example of FAST FAID Model Scores

Date	Time	**FAST**	**FAID**
4/11/2011	14:30	96.26	32
	15:00	96.43	31
	15:30	96.78	31
	16:00	97.26	32
	16:30	97.84	31
	17:00	98.46	29
	17:30	99.06	30
	18:00	99.56	30
	18:30	99.9	30
	19:00	100.01	31
	19:30	99.84	32
	20:00	99.33	34
	20:30	98.45	38
	21:00	97.2	42
	21:30	95.58	44
	22:00	93.63	46
	22:30	91.39	47
	23:00	88.93	48
	23:30	86.32	48
4/12/2011	0:00	83.66	49

The scores for FAST and FAID were arranged in 30 intervals and paired to so that the scores were paired for the same 30 minute interval. These data were then entered into a statistical package and a correlation coefficient was generated. Based on 10,934 FAST-FAID pairs, representing five or six day work schedules, the following statistics were generated.

Descriptive Statistics

	Mean	Std. Deviation	N
FAST	90.63	9.07	10934
FAID	50.07	17.46	10934

There were not an exact number of FAST and FAID scores. The FAID program provides scores at the start, end and for each intervening hour of the work schedule. The FAST program simply calculated the average FAST score for the 30 minute period leading up to the time of day that the work day ended. *InterDynamics* arranged a special run of FAID to produce scores on every half hour of a work schedule and not on the start and end. This enabled half-hourly pairs of FAST and FAID scores to be produced and compared. For the present data set a total of 10795 FAST-FAID pairs were produced and analyzed.

Correlation Between FAST and FAID

		FAST
FAID	Pearson Correlation	-.729(**)
	Sig. (2-tailed)	.000
	N	10795

** Correlation is significant at the 0.01 level (2-tailed).

The bivariate correlation coefficient that was generated from these paired FAST-FAID scores is shown above. The correlation is statistically significant at beyond the P<.001 level and account for 53% of the explained variance. Note that the correlation is negative as would be expected as the FAST scores are higher for lower levels of fatigue while the FAID scores are lower for lower levels of fatigue. The correlation alone indicates the presence of a highly similar relationship between FAST and FAID. There should be no difficulty whatsoever in describing the statistical relationship between the models.

Prediction of FAST Scores

The FRA published a report (Hursh, Raslear, Kaye, and Fanzone, (2006) showing that there is a greater likelihood of human factors caused accidents among persons whose work schedules produce FAST scores below 70. Accordingly, the FRA has accepted FAST as an acceptable method for determining the risk of fatigue in work schedules. Additionally, FRA demonstrated that there was a significant relationship between FAST and FAID scores in its publication (Tabac & Raslear, 2010). Thus, the FRA study suggests that FAST scores below 70 may be sensitive to detecting human factors caused accidents. Other models, like FAID, if they are highly correlated with FAST, can be assumed to show a similar relationship. The goal of a calibration study is to show that the two models are in fact related mathematically. The two data sets obtained were subjected to further analysis using the SPSS Curve Fitting Procedure (SPSS Release 17.0, 2008).

This procedure attempts to fit various mathematical equations to the observed data to estimate the underlying relationship. By understanding the underlying relationship and plotting the data we are able to translate the scores of one model or measuring system to another just as we can convert Fahrenheit to Centigrade on a temperature thermometer. Unfortunately, the two models are not measuring exactly the same thing so we expect that there will not be a perfect translation of the two approaches. Nevertheless, as can be seen, with a correlation of -.73 we have a very high degree of confidence that the models are in fact highly correlated. Nunnaly (1978, p245) describes correlations in the .70 neighborhood as being fairly strong but not describing identical tools.

The table below shows the results for the analysis of the fit to the data of the various curves estimated.

Predicting FAST from FAID: Model Summary and Parameter Estimates

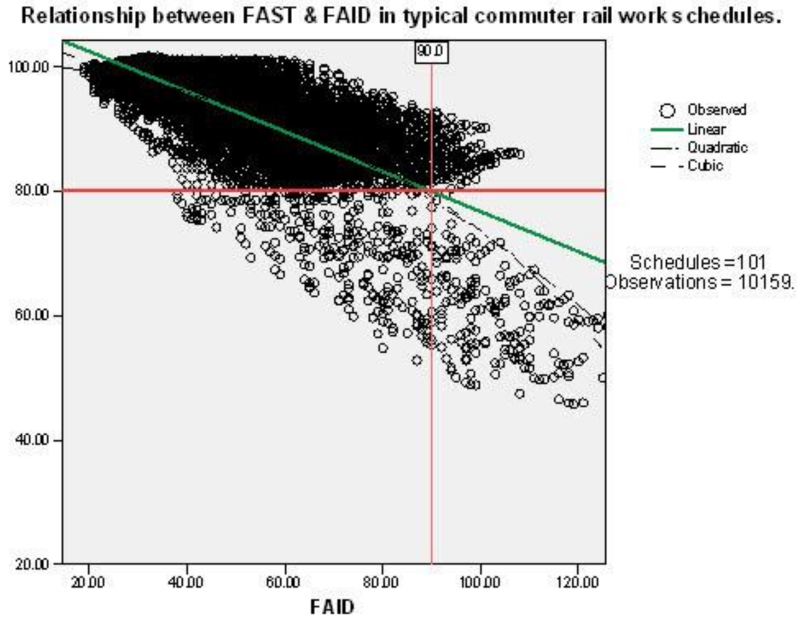
Dependent Variable: FAST

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.531	12239.105	1	10793	.000	108.823	-.321		
Quadratic	.556	6758.315	2	10792	.000	100.461	-.007	-.003	
Cubic	.559	4560.055	3	10791	.000	106.917	-.363	.003	-2.98E-005

The independent variable is FAID.

All of the equations are highly statistically significant in terms of explaining the FAST and FAID scores. The models can be compared to each other by examining the amount of variance accounted for, which is represented in the second column of the table under R Square. This statistic shows how well the model or equation accounts for the observed data. The best model in this case is the Cubic model which accounts for .559 or 56% of the variance, as compared to the others. However, they are all in the same neighborhood and we could not really say at this point that one is highly superior. The cubic is 2.7% better at accounting for the variance and so gets the numerical edge. Most likely the underlying relationship between the models is not linear, but curvilinear. This means that instead of a perfectly straight line, the data are likely arranged in more of a curve with the ends or tails sloping up at either end of the distribution.

As can be seen from the figure below, the purpose of generating the appropriate model is to be able to predict or convert the scores of one model to another. The diagonal line through the center of the darkened section of the graph shows the plot of the relationship between FAST and FAID. The diagonal line is the linear estimate of the relationship between the two models. The vertical red line in the diagram indicates the position of 90 on the FAID scale (which is the current cutoff recommended by the developers of FAID) intersects with the horizontal line from the point of 80 on the FAST axis. The FAID score of 90 corresponds to a score of 80 on the FAST model. The recommended cutoff for FAST is 70. Thus, the present analysis actually identifies a more restrictive or conservative threshold for fatigue than is currently recommended by the FAID authors and than is currently utilized if we accept that the FAST score of 70 as the validated score below which the risk of human factors caused accidents are likely to occur. In other words, by accepting the FAID cutoff of 90 these data suggest that we would obtain a score of 80 on the FAST.

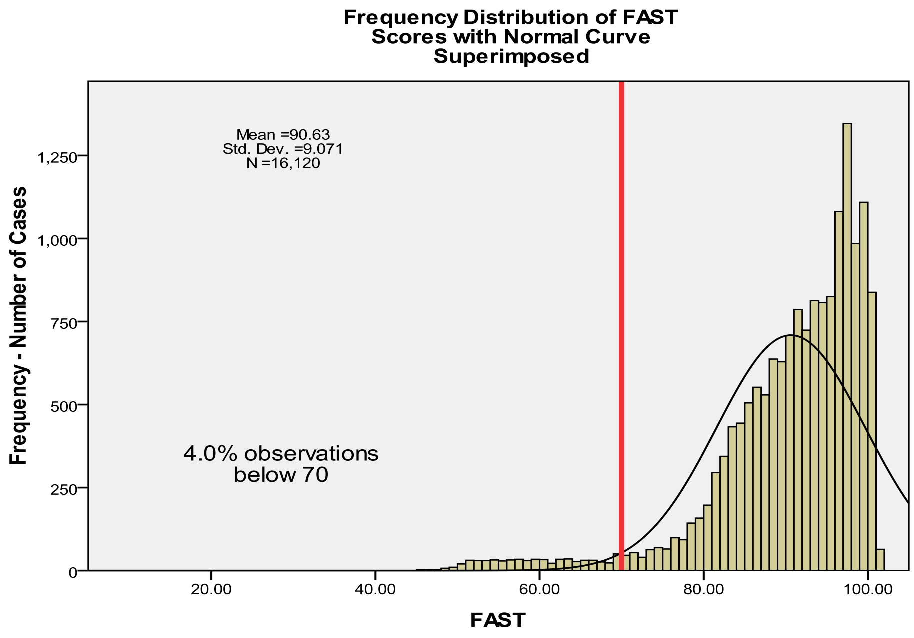

Relationship between FAST & FAID in typical commuter rail work schedules.

To summarize, work schedules which are above 90 on the FAID scale that would be considered to be at risk for fatigue would also be considered at risk for fatigue on the FAST model as well.

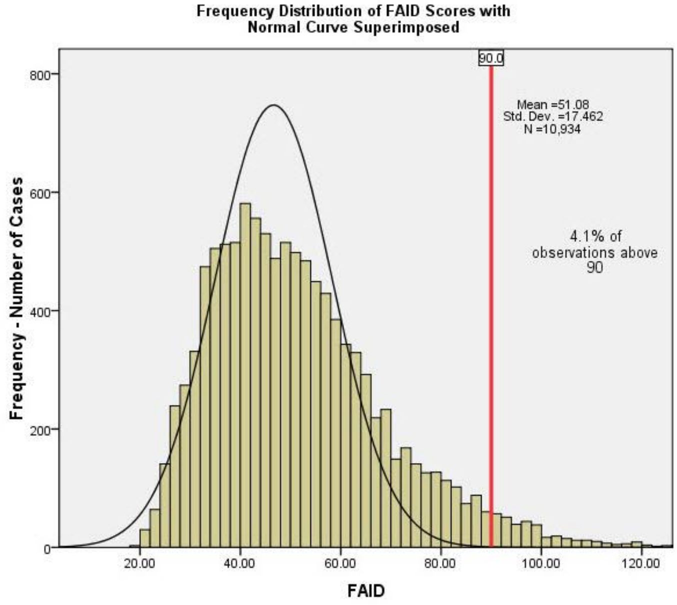
The following histograms present the number of schedule observations that would be identified as at risk for fatigue with varying cutoff scores. For the FAST model, the number of corresponding cases if the fatigue threshold is set at 70 on the FAST model is also 4%.
Using the FAID model if the cutoff were set at 90 (as is recommended) the number of observations falling above the threshold is 4.1%.

The differences between the present study and the Tabac & Raslear (2010) study are likely due to differences between the two samples obtained. In the present study, steps were taken to ensure that a representative sample of work schedules was obtained. The present sample consisted of 61% morning, 36% afternoon and 3% midnight shifts. Thus, the present sample is more reflective of actual work practices as opposed to the more atypical schedules that might have been obtained in the earlier validation study sample (Hursh, Raslear, Kaye, and Fanzone (2006) that was based on human factors and non-human factors caused accidents. Since accidents are so rare in the industry it is clear that theirs was an unusual data set. It should also be noted that the validation sample was obtained entirely from freight operations. The present sample is obtained entirely from commuter rail operations. Examining the means for FAST and FAID reported in the Tabac & Raslear (2010) study on page 16, the mean of FAST and FAID is 69 and 59 respectively. In the present study, FAST and FAID means are 90 and 51 respectively. Thus, the present data is probably more representative of normal working hours and times of day.

FAST Histogram



FAID Histogram



Conclusion

The results indicate the presence of a highly statistically significant relationship between the two models. In addition, the underlying relationships between FAST and FAID are robust and permit the calculation of scores from one model to the other. By being able to compute FAST scores from the FAID model we can assume that the two models are both measuring similar phenomena. Therefore we can assume that the FAID model is also validated. Cutoff scores on the FAID model can also be equated to cutoffs on FAST. Thus, the present analyses indicate that the two models are highly correlated and both reflect the degree of fatigue in the work schedule. The exact score conversion between FAST and FAID is listed below using the linear conversion model.

FAST = (-0.32) *FAID + 108.82

FAST	FAID
70.42	120
72.02	115
73.62	110
75.22	105
76.82	100
78.42	95
80.02	90
81.62	85
83.22	80
84.82	75
86.42	70
88.02	65

Based on these results the evidence suggests that have two tools which measure fatigue. Further, the present evidence suggests that there is no reason to set the FAID cutoff lower than 90.

One cautionary note, these scores are estimates of cognitive effectiveness or readiness to perform tasks at a particular point in time given assumptions that the individual has obtained a reasonable amount of sleep prior to doing so. However, these estimates are based on group averages and are not accurate for the estimate of actual individual performance. Variations of work activity, sleeping schedules and opportunities, not to mention individual differences, will all play a significant role in determining actual readiness. The fatigue models are the best estimate of what we might expect in a certain very general situation. Additional research is needed to improve the accuracy of these models in the workplace.

References

Hursh, S.R., Raslear, T.G., Kaye, A.S., & Fanzone, J.F. (2006). *Validation and calibration of a fatigue assessment tool for railroad work schedules, final report* (Report No. DOT/FRA/ORD-08/04). Washington, DC: U.S. Department of Transportation.

Hursh, S.R., Raslear, T.G., Kaye, A.S., & Fanzone, J.F. (2006). *Validation and calibration of a fatigue assessment tool for railroad work schedules, final report* (Report No. DOT/FRA/ORD-08/04). Washington, DC: U.S. Department of Transportation. (http://www.fra.dot.gov/downloads/Research/ord0804.pdf)

Hursh, S.R., Redmond, D.P., Johnson, M.L. Thorne, D.R., Belenky, G., Balkin, T.J., Storm, W.F., Miller, J.C., & Eddy, D.R. (2004). Fatigue models for applied research in war fighting. *Aviation, Space, and Environmental Medicine*, *75*, A44–A53.

Neri, D.F. (2004). Fatigue and performance modeling workshop, June 13–14, 2002. *Aviation, Space, and Environmental Medicine*, *75*, A1–A3.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Roach, G.D., Fletcher, A., & Dawson, D. (2004).A model to predict work-related fatigue based on hours of work. *Aviation, Space, and Environmental Medicine*, *75*, A61–A69.

Tabak, B, & Raslear, T.G. (2010). Procedures for Validation and Calibration of Human Fatigue Models: The Fatigue Audit InterDyne Tool. (Report No. DOT/FRA/ORD-10/14). Washington, DC: U.S. Department of Transportation.